

PAGE: 1

**RAW SEQUENCE LISTING
PATENT APPLICATION US/08/409,122**DATE: 01/30/97
TIME: 09:39:14**INPUT SET: S15202.raw**

This Raw Listing contains the General
Information Section and up to the first 5 pages.

1 SEQUENCE LISTING
2
3 (1) General Information
4
5 (i) APPLICANT: JOYCE, JAMES G.
6 GEORGE, HUGH A.
7 HOFMANN, KATHRYN J.
8 JANSEN, KATHRIN U.
9 NEEPER, MICHAEL P.
10
11 (ii) TITLE OF THE INVENTION: DNA ENCODING HUMAN PAPILLOMAVIRUS TYPE 18 VACCI
12
13 (iii) NUMBER OF SEQUENCES: 16
14
15 (iv) CORRESPONDENCE ADDRESS:
16 (A) ADDRESSEE: CHRISTINE E. CARTY - MERCK & CO., INC.
17 (B) STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
18 (C) CITY: RAHWAY
19 (D) STATE: NJ
20 (E) COUNTRY: US
21 (F) ZIP: 07065-0907
22
23 (v) COMPUTER READABLE FORM:
24 (A) MEDIUM TYPE: Diskette
25 (B) COMPUTER: IBM Compatible
26 (C) OPERATING SYSTEM: DOS
27 (D) SOFTWARE: FastSEQ Version 1.5
28
29 (vi) CURRENT APPLICATION DATA:
30 (A) APPLICATION NUMBER:US/08/409,122
31 (B) FILING DATE:
32 (C) CLASSIFICATION:
33
34 (vii) PRIOR APPLICATION DATA:
35 (A) APPLICATION NUMBER: 08/408,669
36 (B) FILING DATE: 22-MAR-1995
37
38
39
40 (viii) ATTORNEY/AGENT INFORMATION:
41 (A) NAME: CARTY, CHRISTINE E
42 (B) REGISTRATION NUMBER: 36,099
43 (C) REFERENCE/DOCKET NUMBER: 19425
44
45 (ix) TELECOMMUNICATION INFORMATION:
46 (A) TELEPHONE: 908-594-6734**ENTERED**

**RAW SEQUENCE LISTING
PATENT APPLICATION US/08/409,122.**

DATE: 01/30/97
TIME: 09:39:16

INPUT SET: SIS202.raw

47 (B) TELEFAX: 908-594-4720
48 (C) TELEX:

49

50

51 (2) INFORMATION FOR SEQ ID NO:1:

52

53 (i) SEQUENCE CHARACTERISTICS:

54 (A) LENGTH: 1524 base pairs
55 (B) TYPE: nucleic acid
56 (C) STRANDEDNESS: single
57 (D) TOPOLOGY: linear

58

59 (ii) MOLECULE TYPE: cDNA
60 (iii) HYPOTHETICAL: NO
61 (iv) ANTI-SENSE: NO
62 (v) FRAGMENT TYPE:
63 (vi) ORIGINAL SOURCE:

64

65 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

66

67 ATGGCTTTGT	GGCGGCCCTAG	TGACAATACC	GTATAACCTTC	CACCTCCTTC	TGTGGCAAGA	60
68 GTTGTAATAA	CTGATGATTA	TGTGACTCGC	ACAAGCATAT	TTTATCATGC	TGGCAGCTCT	120
69 AGATTATTAA	CTGTTGGTAA	TCCATATTTT	AGGGTCCCTG	CAGGTGGTGG	CAATAAGCAG	180
70 GATATTCCCA	AGGTTCTGC	ATACCAATAT	AGAGTATTTC	GGGTGCAGTT	ACCTGACCCA	240
71 AATAAAATTG	GTTTACCTGA	TAATAGTATT	TATAATCCTG	AAACACAACG	TTTAGTGTGG	300
72 GCCTGTGCTG	GAGTGGAAAT	TGGCCGTGGT	CAGCCTTCTAG	GTGTTGGCCT	TAGTGGGCAT	360
73 CCATTTTATA	ATAAAATTAGA	TGACACTGAA	AGTTCCCCTG	CCGCTACGTC	TAATGTTCT	420
74 GAGGACGTTA	GGGACAATGT	GTCTGTAGAT	TATAAGCAGA	CACAGTTATG	TATTTGGGC	480
75 TGTGCCCTG	CTATTGGGGA	ACACTGGGCT	AAAGGCACTG	CTTGTAAATC	GCGTCCTTTA	540
76 TCACAGGGCG	ATTGCCCCCC	TTTAACTT	AAGAACACAG	TTTTGGAAGA	TGGTGATATG	600
77 GTAGATACTG	GATATGGTGC	CATGGACTTT	AGTACATTGC	AAGATACTAA	ATGTGAGGTA	660
78 CCATTGGATA	TTTGTCAAGTC	TATTTGTAAA	TATCCTGATT	ATTACAAAT	GTCTGCAGAT	720
79 CCTTATGGGG	ATTCCATGTT	TTTTGCTTA	CGACGTGAGC	AGCTTTTTGC	TAGGCATTTT	780
80 TGGAAATAGGG	CAGGTACTAT	GGGTGACACT	GTGCCTCAAT	CCTTATATAT	TAAAGGCACA	840
81 GGTATGCGTG	CTTCACCTGG	CAGCTGTGTG	TATTCTCCCT	CTCCAAGTGG	CTCTATTGTT	900
82 ACCTCTGACT	CCCAGTTGTT	TAATAAACCA	TATTGGTTAC	ATAAGGCACA	GGGTCTATAAC	960
83 AATGGTATCT	GCTGGCATAA	TCAATTATTT	GTTACTGTGG	TAGATACCAAC	TCGTAGTACC	1020
84 AATTAAACAA	TATGTGCTTC	TACACAGTCT	CCTGTACCTG	GGCAATATGA	TGCTACCAAA	1080
85 TTTAAGCAGT	ATAGCAGACA	TGTTGAAGAA	TATGATTGTC	AGTTTATTTT	TCAGTTATGT	1140
86 ACTATTACTT	TAACTGCAGA	TGTTATGTCC	TATATTCTATA	GTATGAATAG	CAGTATTTA	1200
87 GAGGATTGGA	ACTTTGGTGT	TCCCCCCCCG	CCAACTACTA	GTGGGGTGG	TACATATCGT	1260
88 TTTGTCACAT	CTGTTGCTAT	TACCTGTCAA	AAGGATGCTG	CACCGCTGA	AAATAAGGAT	1320
89 CCCTATGATA	AGTTAAAGTT	TTGGAATGTG	GATTTAAAGG	AAAAGTTTC	TTTGGACTTA	1380
90 GATCAATATC	CCCTTGGACG	TAAATTTTG	GTTCAGGCTG	GATTGCGTGC	CAAGCCCACC	1440
91 ATAGGCCCTC	GTAAACGTTC	TGCTCCATCT	GCCACTACGT	CTTCTAAACC	TGCCAAGCGT	1500
92 GTGCGTGTAC	GTGCCAGGAA	GTAA				1524

93

94 (2) INFORMATION FOR SEQ ID NO:2:

95

96 (i) SEQUENCE CHARACTERISTICS:

97 (A) LENGTH: 507 amino acids
98 (B) TYPE: amino acid
99 (C) STRANDEDNESS: single

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/499,122DATE: 01/30/97
TIME: 09:39:18

INPUT SET: S15202.raw

100 (D) TOPOLOGY: linear
101
102 (ii) MOLECULE TYPE: protein
103 (iii) HYPOTHETICAL: NO
104 (iv) ANTI-SENSE: NO
105 (v) FRAGMENT TYPE: N-terminal
106 (vi) ORIGINAL SOURCE:
107
108 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
109
110 Met Ala Leu Trp Arg Pro Ser Asp Asn Thr Val Tyr Leu Pro Pro Pro
111 1 5 10 15
112 Ser Val Ala Arg Val Val Asn Thr Asp Asp Tyr Val Thr Arg Thr Ser
113 20 25 30
114 Ile Phe Tyr His Ala Gly Ser Ser Arg Leu Leu Thr Val Gly Asn Pro
115 35 40 45
116 Tyr Phe Arg Val Pro Ala Gly Gly Asn Lys Gln Asp Ile Pro Lys
117 50 55 60
118 Val Ser Ala Tyr Gln Tyr Arg Val Phe Arg Val Gln Leu Pro Asp Pro
119 65 70 75 80
120 Asn Lys Phe Gly Leu Pro Asp Asn Ser Ile Tyr Asn Pro Glu Thr Gln
121 85 90 95
122 Arg Leu Val Trp Ala Cys Ala Gly Val Glu Ile Gly Arg Gly Gln Pro
123 100 105 110
124 Leu Gly Val Gly Leu Ser Gly His Pro Phe Tyr Asn Lys Leu Asp Asp
125 115 120 125
126 Thr Glu Ser Ser His Ala Ala Thr Ser Asn Val Ser Glu Asp Val Arg
127 130 135 140
128 Asp Asn Val Ser Val Asp Tyr Lys Gln Thr Gln Leu Cys Ile Leu Gly
129 145 150 155 160
130 Cys Ala Pro Ala Ile Gly Glu His Trp Ala Lys Gly Thr Ala Cys Lys
131 165 170 175
132 Ser Arg Pro Leu Ser Gln Gly Asp Cys Pro Pro Leu Glu Leu Lys Asn
133 180 185 190
134 Thr Val Leu Glu Asp Gly Asp Met Val Asp Thr Gly Tyr Gly Ala Met
135 195 200 205
136 Asp Phe Ser Thr Leu Gln Asp Thr Lys Cys Glu Val Pro Leu Asp Ile
137 210 215 220
138 Cys Gln Ser Ile Cys Lys Tyr Pro Asp Tyr Leu Gln Met Ser Ala Asp
139 225 230 235 240
140 Pro Tyr Gly Asp Ser Met Phe Phe Cys Leu Arg Arg Glu Gln Leu Phe
141 245 250 255
142 Ala Arg His Phe Trp Asn Arg Ala Gly Thr Met Gly Asp Thr Val Pro
143 260 265 270
144 Gln Ser Leu Tyr Ile Lys Gly Thr Gly Met Arg Ala Ser Pro Gly Ser
145 275 280 285
146 Cys Val Tyr Ser Pro Ser Pro Ser Gly Ser Ile Val Thr Ser Asp Ser
147 290 295 300
148 Gln Leu Phe Asn Lys Pro Tyr Trp Leu His Lys Ala Gln Gly His Asn
149 305 310 315 320
150 Asn Gly Ile Cys Trp His Asn Gln Leu Phe Val Thr Val Val Asp Thr
151 325 330 335
152 Thr Arg Ser Thr Asn Leu Thr Ile Cys Ala Ser Thr Gln Ser Pro Val

**RAW SEQUENCE LISTING
PATENT APPLICATION: US/08/409,122**

DATE: 01/30/97

TIME: 09:39:20

INPUT SET: S15202.raw

153	340	345	350
154	Pro Gly Gln Tyr Asp Ala Thr Lys Phe Lys Gln Tyr Ser Arg His Val		
155	355	360	365
156	Glu Glu Tyr Asp Leu Gln Phe Ile Phe Gln Leu Cys Thr Ile Thr Leu		
157	370	375	380
158	Thr Ala Asp Val Met Ser Tyr Ile His Ser Met Asn Ser Ser Ile Leu		
159	385	390	395
160	Glu Asp Trp Asn Phe Gly Val Pro Pro Pro Thr Thr Ser Leu Val		400
161	405	410	415
162	Asp Thr Tyr Arg Phe Val Gln Ser Val Ala Ile Thr Cys Gln Lys Asp		
163	420	425	430
164	Ala Ala Pro Ala Glu Asn Lys Asp Pro Tyr Asp Lys Leu Lys Phe Trp		
165	435	440	445
166	Asn Val Asp Leu Lys Glu Lys Phe Ser Leu Asp Leu Asp Gln Tyr Pro		
167	450	455	460
168	Leu Gly Arg Lys Phe Leu Val Gln Ala Gly Leu Arg Arg Lys Pro Thr		
169	465	470	475
170	Ile Gly Pro Arg Lys Arg Ser Ala Pro Ser Ala Thr Thr Ser Ser Lys		480
171	485	490	495
172	Pro Ala Lys Arg Val Arg Val Arg Ala Arg Lys		
173	500	505	

174

(2) INFORMATION FOR SEQ ID NO:3:

175

176

(i) SEQUENCE CHARACTERISTICS:

177

- (A) LENGTH: 1389 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

178

(ii) MOLECULE TYPE: cDNA

179

(iii) HYPOTHETICAL: NO

180

(iv) ANTI-SENSE: NO

181

(v) FRAGMENT TYPE:

182

(vi) ORIGINAL SOURCE:

183

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

184

185

186

187

188

189

190

191	ATGGTATCCC ACCGTGCCGC ACGACGCAAA CGGGCTTCGG TGACTGACTT ATATAAAACA	60
192	TGTAAACAAT CTGGTACATG TCCATCTGAT GTTGTTAATA AGGTAGAGGG CACCACGTTA	120
193	GCAGATAAAA TATTGCAATG GTCAAGCCTT GGTATATTAA TGGGTGGACT TGGCATAGGT	180
194	ACTGGAAGTG GTACAGGGGG TCGTACAGGG TACATTCCAT TGGGTGGCC TTCCAATACA	240
195	GTTGTGGATG TCGGTCTTAC ACGTCCTCCA GTGGTTATTG AACCTGTGGG CCCCCACAGAC	300
196	CCATCTATTG TTACATTAAT AGAGGACTCA AGTGGTGTAA CATCAGGTGC ACCTAGGCCT	360
197	ACTTTTACTG GCACGTCCTGG GTTTGATATA ACATCTGCTG GTACAACCTAC ACCTGCAGTT	420
198	TTGGATATCA CACCTTCGTC TACCTCTGTT TCTATTCTCA CAACCAATT TACCAATCCT	480
199	GCATTTCTG ATCCGTCAT TATTGAAGTT CCACAAACTG GGGAGGTGTC AGGTAATGTA	540
200	TTTGTGTTGGTA CCCCTACATC TGGAAACACAT GGGTATGAAAG AAATACCTTT ACAAAACATT	600
201	GCTTCTTCTG GTACGGGGGA GGAACCCATT AGTAGTACCC CATTGCCTAC TGTGCGGGCGT	660
202	GTAGCAGGTC CCCGCCTTAA CAGTAGGGCC TACCAACAAG TGTCTGTGGC TAACCTGTGAG	720
203	TTTCTTACAC GTCCCATCCTC TTTAATTACC TATGACAACC CGGCCTTTGA GCCTGTGGAC	780
204	ACTACATTAA CATTGAGGCC TCGTAGTAAT GTTCCTGATT CAGATTTTAT GGATATTATC	840
205	CGTTTACATA GGCCTGCTTT AACATCCAGG CGTGGTACTG TGCGCTTTAG TAGATTAGGT	900

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/409,122

DATE: 01/30/97
 TIME: 09:39:23

INPUT SET: S15202.raw

206	CAAAGGGCAA	CTATGTTTAC	CCGTAGCGGT	ACACAAATAG	GTGCTAGGGT	TCACTTTAT	960
207	CATGATATAA	GTCCTATTGC	ACCCTCCCCA	GAATATATTG	AACTGCAGCC	TTTAGTATCT	1020
208	GCCACGGAGG	ACAATGGCTT	GTTTGATATA	TATGCAGATG	ACATAGACCC	TGCAATGCCT	1080
209	GTACCATCGC	GTCCTACTAC	CTCCTCTGCA	GTTTCTACAT	ATTGCACCCAC	TATATCATCT	1140
210	GCCTCTTCCT	ATAGTAATGT	AACGGTCCCT	TTAACCTCCT	CTTGGGATGT	GCCTGTATAC	1200
211	ACGGGTCTTG	ATATTACATT	ACCACCTACT	ACCTCTGTAT	GGCCCATTGT	ATCACCCACA	1260
212	GCCCCCTGCCT	CTACACAGTA	TATTGGTATA	CATGGTACAC	ATTATTATT	GTGGCCATTA	1320
213	TATTATTTA	TTCCTAAAAA	GCGTAAACGT	GTTCCTATT	TTTTGCAGA	TGGCTTTGTG	1380
214	GGGCCTAG						1389

215

(2) INFORMATION FOR SEQ ID NO:4:

216

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 461 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

217

(ii) MOLECULE TYPE: peptide
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (v) FRAGMENT TYPE: N-terminal
 (vi) ORIGINAL SOURCE:

218

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

219

220	Met Val Ser His Arg Ala Ala Arg Arg Lys Arg Ala Ser Val Thr Asp						
221	1	5	10	15			
222	Leu Tyr Lys Thr Cys Lys Gln Ser Gly Thr Cys Pro Ser Asp Val Val						
223	20	25	30				
224	Asn Lys Val Glu Gly Thr Thr Leu Ala Asp Lys Ile Leu Gln Trp Ser						
225	35	40	45				
226	Ser Leu Gly Ile Phe Leu Gly Gly Leu Gly Ile Gly Thr Gly Ser Gly						
227	50	55	60				
228	Thr Gly Gly Arg Thr Gly Tyr Ile Pro Leu Gly Gly Arg Ser Asn Thr						
229	65	70	75	80			
230	Val Val Asp Val Gly Pro Thr Arg Pro Pro Val Val Ile Glu Pro Val						
231	85	90	95				
232	Gly Pro Thr Asp Pro Ser Ile Val Thr Leu Ile Glu Asp Ser Ser Val						
233	100	105	110				
234	Val Thr Ser Gly Ala Pro Arg Pro Thr Phe Thr Gly Thr Ser Gly Phe						
235	115	120	125				
236	Asp Ile Thr Ser Ala Gly Thr Thr Pro Ala Val Leu Asp Ile Thr						
237	130	135	140				
238	Pro Ser Ser Thr Ser Val Ser Ile Ser Thr Thr Asn Phe Thr Asn Pro						
239	145	150	155	160			
240	Ala Phe Ser Asp Pro Ser Ile Ile Glu Val Pro Gln Thr Gly Glu Val						
241	165	170	175				
242	Ser Gly Asn Val Phe Val Gly Thr Pro Thr Ser Gly Thr His Gly Tyr						
243	180	185	190				
244	Glu Glu Ile Pro Leu Gln Thr Phe Ala Ser Ser Gly Thr Gly Glu Glu						
245	195	200	205				
246	Pro Ile Ser Ser Thr Pro Leu Pro Thr Val Arg Arg Val Ala Gly Pro						

247

248

249

250

251

252

253

254

255

256

257

258

PAGE: 1

**SEQUENCE VERIFICATION REPORT
PATENT APPLICATION US/08/409,122**

DATE: 01/30/97

TIME: 09:39:25

INPUT SET: S15202.raw

Line

Error

Original Text